What is claimed is:

1. A computer-implemented meta search engine method, comprising the steps of:

forwarding a query to a plurality of third party search engines;

parsing the responses from the third party search engines in order to extract information regarding the documents matching the query;

downloading the full text of the documents matching the query;

locating query terms in the documents and extracting text surrounding the query terms; and

displaying the text surrounding the query terms.

2. A method according to Claim 1, further including the step of progressively displaying the text surrounding the query terms as the documents are retrieved.

3. A method according to Claim 1, further including the step of filtering the context strings in order to improve readability by removing redundant whitespace, repeated characters, HTML comments and tags, and special characters.

4. A method according to Claim 1, further including the step of identifying and filtering pages which no longer contain the query terms.

5.   A method according to Claim 1, further including the step of clustering the documents based on analysis of the full text of each document and identification of co-occurring phrases and words, and conjunctions thereof.

6.   A method according to Claim 1, further including the steps of storing the documents matching a query so that a query can be repeated and only showing documents which are new or have been modified since the last query or a given time.

7.   A method according to Claim 1, further including the step of filtering the actual documents when viewed in full in order to (a) highlight the query terms, and (b) insert quick jump links so the user can quickly jump to the query term of interest.

8.   A method according to Claim 1, further including the steps of creating and using a database of meta-information regarding query terms, e.g. storing a list of movie titles, recognizing when the user enters a query containing a movie title, and taking a special action such as referring the user to the review of the movie at a specific movie review site.

9.   A method according to Claim 1, further including the step of storing and using information regarding the particular documents requested by a user in response to a query, e.g. remembering the most commonly requested

document for a given query and presenting this document first in response to the same query in the future.

10. A method according to Claim 1, further including the steps of analyzing the number of documents which would have been found as a function of the number of third party search engines queried, and computing the estimated size of the third party search engines and the estimated size of the document base which the third party search engines index.

11. A method according to Claim 1, further including the step of scheduling regular searches, whereby the user is informed of either new or modified documents since the previous search.

12. A method according to Claim 1, further including the step of using a more advanced detection of duplicate documents by identifying duplicate context even when documents may have different headers or footers.

13. A method according to Claim 1, further including the step of caching the full documents in order to improve access speed.

14. A method according to Claim 1, further including the step of using context sensitive suggestions based on the query entered, e.g. providing suggestions regarding how to search for a name when the query contains a single character that could represent an initial.

-47-

15. A method according to Claim 1, further including the step of using a proximity based ranking scheme to re-rank documents according to the number of and proximity between query terms.

16. A computer-implemented meta search engine method, comprising the steps of:

  forwarding a query to a third party search engine;

  parsing the responses from the third party search engine in order to extract information regarding the documents matching the query;

  downloading the full text of the documents matching the query;

  locating query terms in the documents and extracting text surrounding the query terms; and

  displaying the text surrounding the query terms.

17. A method according to Claim 16, further including the step of progressively displaying the text surrounding the query terms as the documents are retrieved.

18. A method according to Claim 16, further including the step of filtering the context strings in order to improve readability by removing redundant whitespace, repeated characters, HTML comments and tags, and special characters.

19. A method according to Claim 16, further including the step of identifying and filtering pages which no longer contain the query terms.

20. A method according to Claim 16, further including the step of clustering the documents based on analysis of the full text of each document and identification of co-occurring phrases and words, and conjunctions thereof.

21. A method according to Claim 16, further including the steps of storing the documents matching a query so that a query can be repeated and only showing documents which are new or have been modified since the last query or a given time.

22. A method according to Claim 16, further including the step of filtering the actual documents when viewed in full in order to (a) highlight the query terms, and (b) insert quick jump links so the user can quickly jump to the query term of interest.

23. A method according to Claim 16, further including the steps of creating and using a database of meta-information regarding query terms, e.g. storing a list of movie titles, recognizing when the user enters a query containing a movie title, and taking a special action such as referring the user to the review of the movie at a specific movie review site.

24. A method according to Claim 16, further including the step of storing and using information regarding the particular documents requested by a user in response to a query, e.g. remembering the most commonly requested document for a given query and presenting this document first in response to the same query in the future.

25. A method according to Claim 16, further including the step of scheduling regular searches, whereby the user is informed of either new or modified documents since the previous search.

26. A method according to Claim 16, further including the step of using a more advanced detection of duplicate documents by identifying duplicate context even when documents may have different headers or footers.

27. A method according to Claim 16, further including the step of caching the full documents in order to improve access speed.

28. A method according to Claim 16, further including the step of using context sensitive suggestions based on the query entered, e.g. providing suggestions regarding how to search for a name when the query contains a single character that could represent an initial.

29. A method according to Claim 16, further including the step of using a proximity based ranking scheme to re-

rank documents according to the number of and proximity
between query terms.

30. A computer-implemented keyword based image search
engine method, comprising the steps of:
     forwarding a query to a plurality of third
party image search engines;
     parsing the responses from the third party
search engines in order to extract information regarding
the images matching the query;
     downloading the images matching the query; and
     displaying thumbnails of the images to the
user.

31. A method according to claim 30, further including
the step of user selectable filtering of the images based
on size, color, or semantic attributes of the images.

32. A method according to claim 30, further including
the step of identifying and filtering commonly used
images on the Web such as the Netscape Now image and
horizontal bars used to separate sections of a document.

33. A method according to claim 30, further including
the step of identifying and filtering similar images.

34. A method according to claim 30, further including
the steps of identifying the type of an image, e.g.
photograph, line drawing, logo, map, cartoon, portrait,

button, chart, or astronomical pictures, and filtering based on the image type.

35. A method according to claim 30, further including the steps of storing the images matching a query so that a query can be repeated, and only showing new images.

36. A method according to claim 30, further including the step of storing the meta-information (e.g. type of image) so that images may be filtered using the meta-information without downloading the image again for new queries.

37. A method according to claim 30, further including the steps of displaying the full image along with the document referring to it if possible, and highlighting of query terms in the document.

38. A computer-implemented keyword based image search engine method, comprising the steps of:

  forwarding the query to a plurality of third party text search engines;

  parsing the responses from the third party search engines in order to extract information regarding the documents matching the query;

  downloading the documents matching the query;

  analyzing the documents and locating images which may match the user query based on the proximity of query terms to image tags or references;

  downloading the images; and

-52-

displaying thumbnails of the images to the user.

39. A method according to claim 38, further including the step of user selectable filtering of the images based on size, color, or semantic attributes of the images.

40. A method according to claim 38, further including the step of identifying and filtering commonly used images on the Web such as the Netscape Now image and horizontal bars used to separate sections of a document.

41. A method according to claim 38, further including the step of identifying and filtering similar images.

42. A method according to claim 38, further including the steps of identifying the type of an image, e.g. photograph, line drawing, logo, map, cartoon, portrait, button, chart, or astronomical pictures, and filtering based on the image type.

43. A method according to claim 38, further including the steps of storing the images matching a query so that a query can be repeated, and only showing new images.

44. A method according to claim 38, further including the step of storing the meta-information (e.g. type of image) so that images may be filtered using the meta-information without downloading the image again for new queries.

-53-

45. A method according to claim 38, further including the steps of displaying the full image along with the document referring to it if possible, and highlighting of query terms in the document.

46. A computer-implemented meta search engine comprising:

    means for forwarding a query to a plurality of third party search engines;

    means for parsing the responses from the third party search engines in order to extract information regarding the documents matching the query;

    means for downloading the full text of the documents matching the query;

    means for locating query terms in the documents and extracting text surrounding the query terms; and

    means for displaying the text surrounding the query terms.

47. A meta search engine according to Claim 46, further including means for the progressive display of the text surrounding the query terms as the documents are retrieved.

48. A meta search engine according to Claim 46, further including means for the filtering of the context strings in order to improve readability by removing redundant whitespace, repeated characters, HTML comments and tags, and special characters.

49. A meta search engine according to Claim 46, further including means for the identification and filtering of pages which no longer contain the query terms.

50. A meta search engine according to Claim 46, further including a mechanism for clustering the documents based on analysis of the full text of each document and identification of co-occurring phrases and words, and conjunctions thereof.

51. A meta search engine according to Claim 46, further including a mechanism for storing the documents matching a query so that a query can be repeated and for only showing documents which are new or have been modified since the last query or a given date.

52. A computer-implemented meta search engine comprising:

      means for forwarding a query to a third party search engine;

      means for parsing the responses from the third party search engine in order to extract information regarding the documents matching the query;

      means for downloading the full text of the documents matching the query;

      means for locating query terms in the documents and extracting text surrounding the query terms; and

      means for displaying the text surrounding the query terms.

53. A meta search engine according to Claim 52, further including means for the progressive display of the text surrounding the query terms as the documents are retrieved.

54. A meta search engine according to Claim 52, further including means for the filtering of the context strings in order to improve readability by removing redundant whitespace, repeated characters, HTML comments and tags, and special characters.

55. A meta search engine according to Claim 52, further including means for the identification and filtering of pages which no longer contain the query terms.

56. A meta search engine according to Claim 52, further including a mechanism for clustering the documents based on analysis of the full text of each document and identification of co-occurring phrases and words, and conjunctions thereof.

57. A meta search engine according to Claim 52, further including a mechanism for storing the documents matching a query so that a query can be repeated and for only showing documents which are new or have been modified since the last query or a given date.

58. A computer-implemented keyword based image search engine system, comprising:

means for forwarding a query to a number of third party image search engines;

means for parsing the responses from the third party search engines in order to extract information regarding the images matching the query;

means for downloading the images matching the query; and

means for displaying thumbnails of the images to the user.

59.   A system according to claim 58, further including means for selectable filtering of the images based on size, color, or semantic attributes of the images.

60.   A system according to claim 58, further including means for identifying and filtering commonly used images on the Web such as the Netscape Now image and horizontal bars used to separate sections of a document.

61.   A system according to claim 58, further including means for identifying and filtering similar images.

62.   A system according to claim 58, further including means for identifying the type of an image, e.g. photograph, line drawing, logo, map, cartoon, portrait, button, chart, or astronomical pictures, and filtering based on the image type.

63. A system according to claim 58, further including means for storing the images matching a query so that a query can be repeated, and only new images are shown.

64. A system according to claim 58, further including means for storing the meta-information (e.g. type of image) so that images may be filtered using the meta-information without downloading the image again for new queries.

65. A system according to claim 58, further including means for displaying the full image along with the document referring to it if possible, and means for highlighting of query terms in the document.

66. A computer-implemented keyword based image search engine, comprising:

means for forwarding the query to a plurality of third party text search engines;

means for parsing the responses from the third party search engines in order to extract information regarding the documents matching the query;

means for downloading the documents matching the query;

means for analyzing the documents and locating images which may match the user query based on the proximity of query terms to image tags or references;

means for downloading the images; and

means for displaying thumbnails of the images to the user.

-58-

67. A system according to claim 66, further including
means for selectable filtering of the images based on
size, color, or semantic attributes of the images.

68. A system according to claim 66, further including
means for identifying and filtering commonly used images
on the Web such as the Netscape Now image and horizontal
bars used to separate sections of a document.

69. A system according to claim 66, further including
means for identifying and filtering similar images.

70. A system according to claim 66, further including
means for identifying the type of an image, e.g.
photograph, line drawing, logo, map, cartoon, portrait,
button, chart, or astronomical pictures, and filtering
based on the image type.

71. A system according to claim 66, further including
means for storing the images matching a query so that a
query can be repeated, and only new images are shown.

72. A system according to claim 66, further including
means for storing the meta-information (e.g. type of
image) so that images may be filtered using the meta-
information without downloading the image again for new
queries.

73. A system according to claim 66, further including
means for displaying the full image along with the

-59-

document referring to it if possible, and means for highlighting of query terms in the document.

74. A computer-implemented method for estimating the relative coverage of third-party search engines which comprises the steps of:

forwarding a set of queries to two third-party search engines;

retrieving the full list of results from each search engine;

retrieving the text of all pages listed by the search engines;

filtering out pages which are unavailable or no longer match the query;

and

comparing the number of remaining pages from each engine.

75. A computer-implemented method for information retrieval which comprises the steps of:

recognizing a query in the form of a question;

transforming the question into a set of one or more specific forms in

which the answer to the question might be expressed; and

searching for the transformed query.

76. The method according to claim 75, wherein the specific expressive forms for each type of question are manually written.

-60-

77. The method according to claim 75, wherein the specific expressive forms for each type of question are learnt by analyzing the context of query terms in the documents which users select from the search method comprising the steps of:

forwarding a query to a plurality of third party search engines;

parsing the responses from the third party search engines in order to extract information regarding the documents matching the query;

downloading the full text of the documents matching the query;

locating query terms in the documents and extracting text surrounding the query terms;

displaying the text surrounding the query terms; and

identifying common forms of the context.

78. A computer-implemented method for query expansion which comprises the steps of:

stemming the query terms;

searching the set of query result pages for commonly occurring morphological variants of the query terms; and

using the commonly occurring morphological variants for query expansion.